

LIBRARY OF THE
UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

510.84

I l 6 r

no. 565-570



The person charging this material is responsible for its return to the library from which it was withdrawn on or before the **Latest Date** stamped below.

**Theft, mutilation, and underlining of books
are reasons for disciplinary action and may
result in dismissal from the University.**

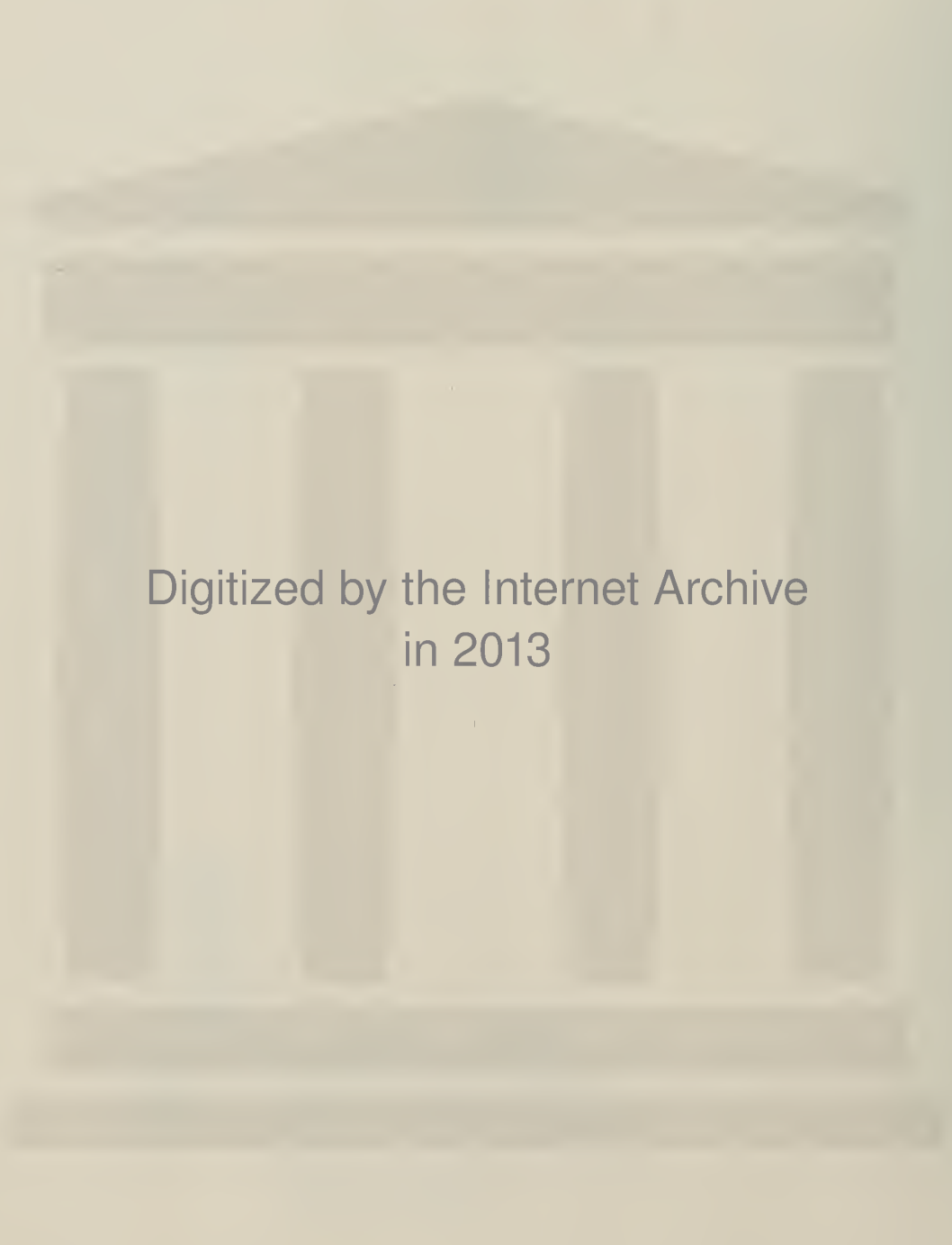
UNIVERSITY OF ILLINOIS LIBRARY AT URBANA-CHAMPAIGN

MAR 22 1978

DEC 14 1976

DEC 12 REC'D

L161—O-1096



Digitized by the Internet Archive
in 2013

<http://archive.org/details/effectofvariable570gear>

34

THE EFFECT OF VARIABLE MESH SIZE ON
THE STABILITY OF MULTISTEP METHODS

by

C. W. Gear
K. W. Tu

MAY 14 1973

April 1973



DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN · URBANA, ILLINOIS

UIUCDCS-R-73-570

THE EFFECT OF VARIABLE MESH SIZE ON
THE STABILITY OF MULTISTEP METHODS*

by

C. W. Gear
K. W. Tu

April 1973

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS 61801

* Supported in part by grant US AEC AT(11-1)1469.

1. INTRODUCTION

This paper is concerned with the stability over the finite interval $t \in [0, T]$ of multistep formulas subject to varying step sizes. Although practical algorithms vary the step size in order to use as large a step as possible consistent with a local control of the estimated error, the existing literature--with the exceptions noted below--deals only with the stability of fixed step formulas.

The standard derivations of multistep formula, e.g. Henrici [3], are based on equal intervals. There are two common ways of handling variable steps for multistep formula. One is to interpolate through the computed points to get values at equally spaced points for use in a standard equal interval formula. The other is to derive multistep formula directly based on unequal intervals. These techniques do not have the same stability characteristics. In 1962 Nordsieck [4] noted an apparent instability in his interpolation version of Adams formula if the step size was varied too frequently. In 1969 Piowtrowski [5] proved that a variable step form of the Adams-Moulton formula was stable and convergent. Brayton, et.al. [1] and Gear [2] (pp. 145-146) point out that the two techniques for changing step size are not equivalent, and Brayton, et.al. give test examples that show that both can cause instability, but that the technique of step variation used by Piotrowski might be more stable.

The stability of a method is affected by three factors: the underlying multistep *formula*, the *technique* used to handle variable steps, and the *scheme* used to select step sizes--the *formula*, *technique*, and *scheme* for short. *Method* will be used to refer to the combination of a formula and a technique. This paper introduces the idea of a step selection scheme, and defines the stability and convergence of a method

with respect to a scheme. A stability condition is derived. It is necessary for stability, and, together with consistency, implies convergence and stability.

It is easy to generate consistent formulas. The more difficult task is to determine whether a combination of formula, technique, and scheme is stable. The two common techniques for step changing are examined, and it is proved that the variable step technique is superior to the interpolation technique for Adams formula. Specifically, it is proved that the variable step technique used with Adams formula is stable and convergent with respect to any step selection scheme, while the interpolation method can be unstable if the steps are allowed to change too much. However, it is proved that the interpolation technique with the k -step Adams formula is stable if the step selection scheme guarantees that the step is constant for at least k -steps after each change.

The investigation is extended to all multistep formulas, whether explicit, implicit, or predictor-corrector. It is proved that if the underlying formula is stable, then both step changing techniques lead to stable methods if the rate of change of the step size is small. It is also shown that common error control algorithms lead to small rates of change in the step size.

2. BACKGROUND

The two step changing techniques will be called *interpolation* and *variable step*, respectively. Intuitively, we define them as follows. Suppose we are using a k -step formula for integration and suppose that an approximation y_{n-i} to the solution is known at k values of the independent variable t , say t_{n-i} , $k > i \geq 0$. The interpolation technique for changing step size to h consists of the following two steps:

1. Interpolate through the known points to get the values of the solution at the points $\tilde{t}_{n-i} = t_n - ih$, $k > i \geq 0$.
2. Use the fixed step integration formula on these new points $\tilde{t}_{n-i}$ to find the value at $t_{n+1} = t_n + h$.

In the variable step technique we start with k unequally spaced points t_{n-i} , $k > i \geq 0$ and compute the coefficients of the multistep formula

$$y_{n+1} = \alpha_{1,n} y_n + \dots + \alpha_{k,n} y_{n-k+1} + \beta_{0,n} h_n y'_{n+1} \quad (1)$$

$$+ \beta_{1,n} h_{n-1} y'_n + \dots + \beta_{k,n} h_{n-k} y'_{n-k+1}$$

so that it is exact to the required order, where $h_n = t_{n+1} - t_n$. Since the order cannot exceed $2[k/2]+2$ if the formula is to be stable for fixed step sizes, additional restrictions must be imposed. These consist of specifying the values of some of the coefficients α and β . The fixed step formula underlying the variable step technique is said to be the formula that is obtained from (1) when equally spaced points t_{n-i} are used. For example, if we required that $\alpha_{j,n} = 0$, $j \geq 2$, and that (1) have order $k+1$, then Adams-Moulton is the underlying fixed step formula. If we also ask that $\beta_{0,n} = 0$ and lower the order to k , we get a variable step method based on the Adams-Bashforth formula.

The stability of an integration formula subject to a step changing technique may depend on the way in which the steps are varied. Therefore, we must define stability with respect to a particular step selection scheme. First, therefore, we state

Definition 1 A step selection scheme is a function θ such that

$$h_n = h\theta(t_n, h)$$

where, for all $h > 0$, $0 \leq t \leq T$

$$1 \geq \theta(t, h) \geq \Delta > 0$$

NOTE: As h is reduced to zero, the maximum step size is reduced to zero.

The lower bound Δ serves to prevent zero step sizes for any non-zero h ; these would not give useful numerical integration schemes!

In this paper we will be concerned only with differential equations $y' = f(y, t)$ for which $f(y, t)$ is Lipschitz continuous in the strip $|y| < \infty$, $0 \leq t \leq T < \infty$. Small perturbations to the solutions of such equations at a given point cause small perturbations later on. Stability means that small numerical disturbances at one point t_n cause small perturbations in the numerical solution at later times for all small h . Convergence means that the numerical solution can be made arbitrarily accurate as $h \rightarrow 0$.

Definition 2 A method is stable with respect to a scheme θ if there exists a constant $M < \infty$ (dependent on the differential equation only) such that

$$|y_m - \hat{y}_m| < M |y_n - \hat{y}_n|$$

for all $0 \leq t_n < t_m \leq T$, where y_i and $\hat{y}_i$ are two numerical solutions.

Definition 3 A method is convergent with respect to a scheme θ if the computed solution y_n converges to $y(t_n)$ for any $0 \leq t_n \leq T$ as $h \rightarrow 0$ and the starting errors tend to zero.

We will only discuss a single differential equation. The extension to systems of equations is straightforward, although notationally complex. We will also assume that $f(y,t)$ and $y(t)$ are differentiable as often as necessary to save stating all the conditions each time. The results are valid if f is only Lipschitz continuous, and can be obtained by replacing $f_y = \partial f / \partial y$ by the Lipschitz constant L whenever f_y appears.

3. NOTATIONAL DEVELOPMENT

By developing a uniform notation for all the formulas and step changing techniques to be discussed, we will be able to give general proofs in Section 4 that can be applied to all cases. Some of this notation has been developed in [2]. It is summarized briefly here and suitably extended.

Consider the integration formula (1) with $\beta_{0,n} = 0$. It is an explicit formula which can be used directly. It can also be used as a predictor for a corrector formula. We will write the corrector formula as

$$y_{n+1}^{(m+1)} = \alpha_{1,n}^* y_n + \dots + \alpha_{k,n}^* y_{n-k+1} + \beta_{0,n}^* h_n f(y_{n+1}^{(m)}, t_{n+1}) \quad (2)$$

$$+ \beta_{1,n}^* h_{n-1} y_n' + \dots + \beta_{k,n}^* h_{n-k} y_{n-k+1}'$$

A predictor-corrector formula is one in which a predicted value $y_{n+1}^{(0)}$ is computed by the explicit equation

$$y_{n+1}^{(0)} = \alpha_{1,n} y_n + \dots + \alpha_{k,n} y_{n-k+1} \quad (3)$$

$$+ \beta_{1,n} h_{n-1} y_n' + \dots + \beta_{k,n} h_{n-k} y_{n-k+1}'$$

and the equation (2) is used to compute a fixed finite number of corrected values $y_{n+1}^{(1)}, \dots, y_{n+1}^{(M)}$. The approximation y_{n+1} is defined as $y_{n+1}^{(M)}$ and y_{n+1}' is set to either $f(y_{n+1}^{(M)})$ or $f(y_{n+1}^{(M-1)})$ depending on whether or not a final function evaluation is needed.

An implicit formula is one in which equation (2) is solved for $y_{n+1}^{(m+1)} = y_n^{(m)}$. One way of doing this is to iterate a predictor-corrector formula until successive iterates are approximately equal. This converges for h sufficiently small but may diverge for large h . Other methods, such as Newton-Raphson, can also be used to solve (2). The stability of a method based on an implicit formula is independent of the process used to solve equation (2), so we do not need to specify that process.

During the computation, a number of previous values of the function y_{n-i} and the scaled derivatives $h_{n-i-1} y'_{n-i}$ must be saved. Let us collect these together in a column vector $\underline{y}_n$ defined by

$$\underline{y}_n = [y_n, y_{n-1}, \dots, y_{n-k+1}, h_{n-1} y'_n, h_{n-2} y'_{n-1}, \dots, h_{n-k} y'_{n-k+1}]^T$$

for a k -step formula, where T is the transpose operator. (If some of these values are not needed, we can drop them from the vector. For example, in Adams formula, $\underline{y}_n$ would be $[y_n, h_{n-1} y'_n, \dots, h_{n-k} y'_{n-k+1}]^T$.)

A numerical integration method is a method for computing $\underline{y}_{n+1}$ from $\underline{y}_n$. We will first describe it for the variable step technique.

Let $\underline{y}_{n+1}^{(m)}$ be defined as

$$[y_{n+1}^{(m)}, y_n, \dots, y_{n-k+2}, h_n y'_{n+1}^{(m)}, h_{n-1} y'_n, \dots, h_{n-k+1} y'_{n-k+2}]^T$$

where $h_n y'_{n+1}^{(m)}$ is $h_n f(y_{n+1}^{(m-1)}, t_{n+1})$ for $m \geq 1$,

$$h_n y'_{n+1}^{(0)} = \gamma_{1,n} y_n + \gamma_{2,n} y_{n-1} + \dots + \gamma_{k,n} y_{n-k+1}$$

$$+ \delta_{1,n} h_{n-1} y'_n + \delta_{2,n} h_{n-2} y'_{n-1} + \dots + \delta_{k,n} h_{n-k} y'_{n-k+1}$$

when $m = 0$, and the coefficients γ and δ are given by

$$\gamma_{i,n} = (\alpha_{i,n} - \alpha_{i,n}^*) / \beta_{0,n}^*$$

and

$$\delta_{i,n} = (\beta_{i,n} - \beta_{i,n}^*) / \beta_{0,n}^*$$

From these definitions we can write

$$\underline{y}_{n+1}^{(0)} = A_n \underline{y}_n \quad (4)$$

where $A_n =$

$$\begin{bmatrix} \alpha_{1,n} & \alpha_{2,n} & \dots & \alpha_{k-1,n} & \alpha_{k,n} & \beta_{1,n} & \beta_{2,n} & \dots & \beta_{k-1,n} & \beta_{k,n} \\ 1 & 0 & \dots & 0 & 0 & & & & & \\ 0 & 1 & \dots & 0 & 0 & & & 0 & & \\ & \dots & \dots & & \dots & & & & & \\ 0 & 0 & \dots & 1 & 0 & & & & & \\ \hline \gamma_{1,n} & \gamma_{2,n} & \dots & \gamma_{k-1,n} & \gamma_{k,n} & \delta_{1,n} & \delta_{2,n} & \dots & \delta_{k-1,n} & \delta_{k,n} \\ & & & & & 1 & 0 & \dots & 0 & 0 \\ & & & 0 & & 0 & 1 & \dots & 0 & 0 \\ & & & & & & \dots & \dots & \dots & \\ & & & & & 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

and

$$y_{n+1}^{(m+1)} = y_{n+1}^{(m)} + \underline{\ell}_n [h_n f(y_{n+1}^{(m)}, t_{n+1}) - h_n y_{n+1}'^{(m)}] \quad (5)$$

where

$$\underline{\ell}_n = [\beta_{0,n}^*, 0, \dots, 0, 1, 0, \dots, 0]^T \quad (6)$$

Equation (4) represents the predictor--the evaluation of $y_{n+1}^{(0)}$ and $h_n y_{n+1}'^{(0)}$ as linear combinations of the known information plus the shifting over of the saved information and the discarding of y_{n-k+1} and y_{n-k+1}' . Equation (5) represents the corrector--the evaluation of $f(y_{n+1}^{(m)}, t_{n+1})$ and the updating of $y_{n+1}^{(m)}$ and $h_n y_{n+1}'^{(m)}$. For convenience, we define

$$F_n(y_{n+1}^{(m)}) = h_n f(y_{n+1}^{(m)}, t_{n+1}) - h_n y_{n+1}'^{(m)} \quad (7)$$

Then equation (5) can be written

$$\underline{y}_{n+1}^{(m+1)} = \underline{y}_{n+1}^{(m)} + \underline{\ell}_n F_n(\underline{y}_{n+1}^{(m)}) \quad (8)$$

If a predictor only formula is used, we still formally must apply one step of (8) with $\underline{\ell}_n = \underline{\ell}^P = [0, \dots, 0, 1, 0, \dots, 0]^T$. (In this case the value of α^* and β^* are unimportant.) If a final function evaluation is used, we apply (8) M times using the $\underline{\ell}_n$ given in (6), and then apply one further step of (8) using $\underline{\ell}_n = \underline{\ell}^P$.

Now let us consider a fixed step method in preparation for formalizing the interpolation technique. In this case, the α , β , γ , and δ in A_n and $\underline{\ell}_n$ are independent of n . Suppose we have just completed the step from t_{n-1} to t_n using step h_{n-1} . The values saved in $\underline{y}_n$ are the computed approximations at the points $t_n - ih_{n-1}$, $0 \leq i < k$. Before stepping to $t_{n+1} = t_n + h_n$ we must interpolate to the points $\tilde{t}_{n-i} = t_n - ih_n$. This is a linear process and can be represented by first premultiplying $\underline{y}_n$ by an interpolating matrix C_n . Consequently, we write the predictor step for the interpolation process as

$$\underline{y}_{n+1}^{(0)} = A C_n \underline{y}_n \quad (9)$$

The corrector step (8) is unchanged except that $\underline{\ell}_n = \underline{\ell}$ is independent of n .

Both for theoretical development and computational ease, it is convenient to perform a transformation based on Nordsieck's [4] form of Adams formula for the interpolation technique. This consists of computing and saving $Q \underline{y}_n$ rather than $\underline{y}_n$ where Q is a nonsingular matrix. The particular form of Q used is obtained as follows: The set of values of y_{n-i} and y'_{n-i} $0 \leq i < k$ determine a unique $(2k-1)$ -th degree polynomial.

Represent that polynomial instead by its value at t_n and the values of its first $2k-1$ scaled derivatives $h_{n-1}^j y^{(j)}/j!$, $1 \leq j < 2k$. (This leads to what is called a $2k$ -value method in Gear [2].) If $\underline{y}_n$ is the set of saved values of y and hy' based on step h_{n-1} , and $\underline{a}_n$ is the set of values of $[y_n, h_{n-1} y'_n, h_{n-1}^2 y''_n/2, \dots, h_{n-1}^{2k-1} y_n^{(2k-1)}/(2k-1)!]^T$ representing the same polynomial of degree $2k-1$, then Q is defined by $\underline{a}_n \equiv Q \underline{y}_n$. Q is independent of y and h_{n-1} . If we apply this transformation to (9) and (8) we get

$$Q \underline{y}_{n+1}^{(0)} = (QAQ^{-1})(QC_n Q^{-1})Q \underline{y}_n \quad (10)$$

$$Q \underline{y}_{n+1}^{(m+1)} = Q \underline{y}_{n+1}^{(m)} + Q \underline{\ell}_n F_n(Q^{-1}(Q \underline{y}_{n+1}^{(m)})) \quad (11)$$

To simplify the presentation we will write $\underline{y}_n$ rather than $\underline{a}_n$, A rather than QAQ^{-1} , C_n rather than $QC_n Q^{-1}$, $\underline{\ell}_n$ rather than $Q\underline{\ell}_n$, and $F(\underline{c})$ rather than $F(Q^{-1}\underline{c})$ when we are talking about the transformed scheme.

Consequently, equations (10) and (11) are identical to (9) and (8).

The advantage of making the transformation is that

$$C_n = \text{diag}[1, h_n/h_{n-1}, \dots, (h_n/h_{n-1})^{2k-1}] \quad (12)$$

so that the interpolation process is simplified. This was Nordsieck's motivation. Finally, we note that (9) is identical to (4) if A_n is interpreted as $A C_n$ for the interpolation technique. Consequently, equations (4) and (8) represent the predictor only formula or the prediction-correction formula in either step changing technique.

The corrector only formula can sometimes be implemented by iterating the corrector to convergence and treated theoretically by setting $\underline{y}_{n+1} = \lim_{m \rightarrow \infty} \underline{y}_{n+1}^{(m)}$. However, this process may not converge (as in stiff equations), in which case the corrector must be solved by other

means. We note that the corrector equation (8) always adds on a multiple of $\frac{\ell}{n}$ to the current value. Consequently, the final value has the form

$$\underline{y}_{n+1} = \underline{y}_{n+1}^{(0)} + \omega \frac{\ell}{n} = A_n \underline{y}_n + \omega \frac{\ell}{n} \quad (13)$$

where ω is a scalar. The value of ω is such that

$$F(\underline{y}_{n+1}^{(0)} + \omega \frac{\ell}{n}) = 0 \quad (14)$$

(Although we have used predictor coefficients in A_n , some straightforward arithmetic will show that the final value of $\underline{y}_{n+1}$ is independent of these coefficients when the corrector equation is solved by (13) and (14).)

We must now develop the error propagation formula for equations (4) and either (8) or (13) and (14). The global error ϵ_n is defined by

$$\epsilon_n = \underline{y}_n - \underline{y}(t_n) \quad (15)$$

where $\underline{y}(t_n)$ is the value of the components of the vector $\underline{y}_n$ as evaluated along the desired solution of the differential equation.

The truncation error $\underline{d}_n$ is the error introduced by one step of the method, starting from the solution of the differential equation. Thus, $\underline{d}_n$ is defined by the equations

$$\begin{aligned} \tilde{\underline{y}}_{n+1}^{(0)} &= A_n \underline{y}(t_n) \\ \tilde{\underline{y}}_{n+1}^{(m+1)} &= \tilde{\underline{y}}_{n+1}^{(m)} + \frac{\ell}{n} F_n(\tilde{\underline{y}}_{n+1}^{(m)}) , \end{aligned}$$

either

$$\tilde{\underline{y}}_{n+1} = \tilde{\underline{y}}_{n+1}^{(M)}$$

if M corrector iterations are used, or

$$\tilde{\underline{y}}_{n+1} = \tilde{\underline{y}}_{n+1}^{(0)} + \tilde{\omega} \frac{\ell}{n} = A_n \underline{y}(t_n) + \tilde{\omega} \frac{\ell}{n} , \quad (16)$$

where $F(\tilde{y}_{n+1}) = 0$ if the corrector is solved exactly, and

$$\underline{d}_n = \tilde{y}_{n+1} - y(t_{n+1})$$

From the definition of $\underline{\varepsilon}_{n+1}$ and $\underline{d}_n$

$$\begin{aligned}\underline{\varepsilon}_{n+1} &= y_{n+1} - y(t_{n+1}) \\ &= y_{n+1} - \tilde{y}_{n+1} + \tilde{y}_{n+1} - y(t_{n+1}) \\ &= y_{n+1} - \tilde{y}_{n+1} + \underline{d}_n\end{aligned}\quad (17)$$

(Note that if round-off errors are to be considered, they can be included in $\underline{d}_n$ at this point, so that the computation of y_{n+1} can be considered to be exact.) If M corrector iterations are used, (17) can be written as

$$\begin{aligned}\underline{\varepsilon}_{n+1} &= y_{n+1}^{(M)} - \tilde{y}_{n+1}^{(M)} + \underline{d}_n \\ &= y_{n+1}^{(M-1)} + \ell_n F_n(y_{n+1}^{(M-1)}) - \tilde{y}_{n+1}^{(M-1)} - \ell_n F_n(\tilde{y}_{n+1}^{(M-1)}) + \underline{d}_n \\ &= [I + \ell_n \frac{\partial F}{\partial y}^{(M)}](y_{n+1}^{(M-1)} - \tilde{y}_{n+1}^{(M-1)}) + \underline{d}_n\end{aligned}\quad (18)$$

by the mean value theorem. $\frac{\partial F}{\partial y}^{(M)}$ is the row vector of partial derivatives of F_n with respect to the components of y evaluated somewhere in the interval $(y_{n+1}^{(M-1)}, \tilde{y}_{n+1}^{(M-1)})$. From equation (7) we see that

$$\frac{\partial F}{\partial y}^{(M)} = [h_n f_y^{(M)}, 0, \dots, 0, -1, 0, \dots, 0]$$

where $f_y^{(M)}$ is $\partial f / \partial y$ evaluated in the interval $(y_{n+1}^{(M-1)}, \tilde{y}_{n+1}^{(M-1)})$. The process leading to (18) can be repeated to get

$$\begin{aligned}\underline{\varepsilon}_{n+1} &= [I + \ell_n \frac{\partial F}{\partial y}^{(M)}] \dots [I + \ell_n \frac{\partial F}{\partial y}^{(1)}] (y_{n+1}^{(0)} - \tilde{y}_{n+1}^{(0)}) + \underline{d}_n \\ &= [I + \ell_n \frac{\partial F}{\partial y}^{(M)}] \dots [I + \ell_n \frac{\partial F}{\partial y}^{(1)}] A_n \underline{\varepsilon}_n + \underline{d}_n\end{aligned}\quad (19)$$

We write this in the form

$$\underline{\varepsilon}_{n+1} = S_n \underline{\varepsilon}_n + \underline{d}_n \quad (20)$$

Let $\underline{e}_i^T$ be the unit row vector consisting of a one in the i -th position and zeros elsewhere. Thus, $\underline{e}_1^T$ is $[1, 0, \dots, 0]$. Let subscript d represent the component corresponding to the $h_{n-1} y_n'$ entry in $\underline{y}_n$. (d is $k+1$ for the untransformed method, 2 for the transformed method.) Then

$$\frac{\partial F^{(m)}}{\partial \underline{y}} = h_n f_y^{(m)} \underline{e}_1^T - \underline{e}_d^T \quad (21)$$

Therefore, from (19), (20), and (21)

$$\begin{aligned} S_n &= \prod_{m=1}^M [(I - \underline{\ell}_n \underline{e}_d^T) + h_n f_y^{(m)} \underline{\ell}_n \underline{e}_1^T] A_n \\ &= [(I - \underline{\ell}_n \underline{e}_d^T)]^M + \Sigma \{ \text{Products of } M \text{ terms of form} \\ &\quad h_n f_y^{(m)} \underline{\ell}_n \underline{e}_1^T \text{ or } (I - \underline{\ell}_n \underline{e}_d^T) \\ &\quad \text{with at least one of the former} \} A_n \end{aligned} \quad (22)$$

We define

$$\hat{S}_n = (I - \underline{\ell}_n \underline{e}_d^T)^M A_n \quad (23)$$

Then we can write (22) as

$$S_n = \hat{S}_n + h_n \tilde{S}_n \quad (24)$$

where $\tilde{S}_n$ is a matrix whose elements are polynomials in h_n , $f_y^{(m)}$ and entries from A_n and $\underline{\ell}_n$.

Lemma 1 If α , β , α^* and β^* are uniformly bounded or if A_n and ℓ_n are uniformly bounded so is $||\tilde{S}_n||$ for all $|h| \leq u$.

Proof The elements of $\tilde{S}_n$ are polynomials in h_n whose coefficients are polynomials in $f_y^{(m)}$ and the entries of A_n and ℓ_n . $|f_y^{(m)}|$ is bounded by the Lipschitz constant, while entries in A_n and ℓ_n , except possibly for the d-th row of A_n , are bounded if α , β , α^* and β^* are bounded. The d-th row of A_n contains elements of the form $(\alpha - \alpha^*)/\beta_{0,n}^*$, and could be unbounded if $\beta_{0,n}^*$ can approach 0. However, from (22) and (24), the only way the d-th row of A_n enters S_n is when A_n is premultiplied by $(I - \ell_n e_d^T)$ which zeros the d-th row and changes the first row from α to α^* . Consequently, the coefficients of h_n in $\tilde{S}_n$ are bounded, so for any $u > 0$, $||\tilde{S}_n||$ is bounded for all $0 \leq h \leq u$. If A_n and ℓ_n are uniformly bounded, the result follows similarly.

Q.E.D.

The behavior of the error ε_n is determined by equation (20). When $||\tilde{S}_n||$ is bounded, we can use equation (24) to relate the stability of the method to the properties of $\hat{S}_n$. The following condition will be shown to be necessary for a method to be stable:

Definition 4 A method satisfies the stability condition with respect to a step selection scheme if there exists a $C < \infty$, such that

$$\hat{S}_n^m = \hat{S}_{m-1} \hat{S}_{m-2} \dots \hat{S}_n \quad (25)$$

satisfies

$$||\hat{S}_n^m|| \leq C \quad (26)$$

for all $0 \leq t_n < t_m \leq T$.

If an implicit method is used, the corrector is solved exactly, and we must derive equations (20) and (24) by a different technique. Substitute (13) and (16) in (17) to get

$$\underline{\varepsilon}_{n+1} = A_n \underline{\varepsilon}_n + (\omega - \tilde{\omega}) \underline{\ell} + \underline{d}_n \quad (27)$$

(The subscript n has been omitted from $\underline{\ell}$ for clarity later.) We know

$$\begin{aligned} 0 &= F(\underline{y}_{n+1}) - F(\underline{\tilde{y}}_{n+1}) \\ &= \frac{\partial F}{\partial \underline{y}} (\underline{y}_{n+1} - \underline{\tilde{y}}_{n+1}) \\ &= \frac{\partial F}{\partial \underline{y}} (A_n \underline{\varepsilon}_n + (\omega - \tilde{\omega}) \underline{\ell}) \end{aligned}$$

Hence,

$$(\omega - \tilde{\omega}) = -\left\{ \frac{\partial F}{\partial \underline{y}} \underline{\ell} \right\}^{-1} \left\{ \frac{\partial F}{\partial \underline{y}} A_n \underline{\varepsilon}_n \right\}$$

Note that both quantities enclosed in braces are scalars. Substituting in (27) we get

$$\begin{aligned} \underline{\varepsilon}_{n+1} &= A_n \underline{\varepsilon}_n - \left\{ \frac{\partial F}{\partial \underline{y}} \underline{\ell} \right\}^{-1} \underline{\ell} \left\{ \frac{\partial F}{\partial \underline{y}} A_n \underline{\varepsilon}_n \right\} + \underline{d}_n \\ &= \left(I - \left\{ \frac{\partial F}{\partial \underline{y}} \underline{\ell} \right\}^{-1} \underline{\ell} \frac{\partial F}{\partial \underline{y}} \right) A_n \underline{\varepsilon}_n + \underline{d}_n \\ &= S_n \underline{\varepsilon}_n + \underline{d}_n \end{aligned}$$

where

$$\begin{aligned} S_n &= [I - \{\frac{\partial F}{\partial \underline{y}} \underline{\ell}\}^{-1} \underline{\ell} \frac{\partial F}{\partial \underline{y}}] A_n \\ &= [I - \{\ell_1 h_n f_y - \ell_d\}^{-1} \underline{\ell} (h_n f_y \underline{e}_1^T - \underline{e}_d^T)] A_n \end{aligned}$$

Here, ℓ_i is the i -th component of $\underline{\ell}$.

If we choose u such that for $0 < h_n \leq u$, $h_n \ell_1 f_y - \ell_d$ is bounded away from zero (ℓ_d is one in the methods discussed), then we can write

$\{h_n \ell_1 f_y - \ell_d\}^{-1}$ as $-\ell_d^{-1} \{1 + \ell_1 h_n f_y / \ell_d + O(h_n^2)\}$. Hence,

$$\begin{aligned} S_n &= [I - \ell_d^{-1} \underline{\ell} \underline{e}_d^T] A_n + h_n \tilde{S}_n \\ &= \hat{S}_n + h_n \tilde{S}_n \end{aligned} \quad (28)$$

The form of $\hat{S}_n$ is the same as the previous form with $M = 1$ and $\underline{\ell}$ scaled so that $\ell_d = 1$, as it always is in methods discussed in this paper.

We also need to extend Lemma 1 to this case.

Lemma 1a $\|\tilde{S}_n\|$ defined in (28) is bounded if α , β , α^* and β^*

or A_n and $\underline{\ell}_n$ are uniformly bounded.

Proof From (28) and the preceding equation

$$\begin{aligned} \tilde{S}_n &= -\{\ell_1 h_n f_y - \ell_d\}^{-1} f_y \underline{\ell} \underline{e}_1^T A_n \\ &\quad + [\{\ell_1 h_n f_y - \ell_d\}^{-1} + \ell_d^{-1}] \underline{\ell} \underline{e}_d^T A_n / h_n \\ &= -\{\ell_1 h_n f_y - \ell_d\}^{-1} f_y \underline{\ell} \underline{e}_1^T A_n \\ &\quad + \ell_d^{-1} \{\ell_1 h_n f_y - \ell_d\}^{-1} f_y \underline{\ell} \ell_1 \underline{e}_d^T A_n \end{aligned}$$

All but possibly the d -th row of A_n are bounded. The first term does not involve the d -th row of A_n , and is, therefore, bounded for $0 \leq h \leq u$. The second term involves the d -th row of A_n , but scales it by $\ell_1 = \beta_{0,n}^*$. Hence, it is also bounded for $0 \leq h \leq u$. Hence, $||\tilde{S}_n||$ is bounded. If A_n and ℓ_n are bounded, the result follows directly.

Q.E.D.

4. STABILITY AND CONVERGENCE

Theorem 1 *If a method is stable with respect to a step selection scheme θ , it satisfies the stability condition (26) with respect to that scheme.*

Proof Consider the differential equation $y' = 0$ for which $\tilde{S}_n$, defined by (24), is zero. Evidently, the difference between two numerical solutions satisfies

$$\underline{y}_{n+1} - \underline{\hat{y}}_{n+1} = \hat{S}_n(\underline{y}_n - \underline{\hat{y}}_n)$$

If the $|\hat{S}_n^m|$ are not uniformly bounded, there exist t_m and t_n and bounded $\underline{y}_n - \underline{\hat{y}}_n$ such that

$$\underline{y}_m - \underline{\hat{y}}_m = \hat{S}_n^m(\underline{y}_n - \underline{\hat{y}}_n)$$

is arbitrarily large, implying that the method is not stable with respect to θ . Hence stability implies the stability condition.
Q.E.D.

The convergence theorem depends on the following lemma.

Lemma 2 If, for the difference equation,

$$\underline{x}_{n+1} = \hat{S}_n \underline{x}_n + h_n \tilde{S}_n \underline{x}_n + h_n \underline{\lambda}_n \quad (29)$$

there exist constants k_0 , k_1 , and k_2 such that

$$||\tilde{S}_n|| \leq k_1 \quad \forall n,$$

$$||\lambda_n|| \leq k_2 \quad \forall n,$$

and $||\hat{S}_n^m|| \leq k_0 \quad \forall m \geq n$

where

$$\hat{S}_n^n = I$$

and $\hat{S}_n^m$ is defined by equation (25),

then, if $k_1 > 0$

$$||x_N|| \leq [k_0 ||x_0|| + \frac{k_2}{k_1}] e^{k_0 k_1 (t_N - t_0)} - \frac{k_2}{k_1} \quad (30)$$

or if $k_1 = 0$

$$||x_N|| \leq k_0 ||x_0|| + k_0 k_2 (t_N - t_0) \quad (31)$$

Proof From (29) we have

$$x_N = \hat{S}_0^N x_0 + \sum_{n=0}^{N-1} \hat{S}_{n+1}^N h_n (\tilde{S}_n x_n + \lambda_n)$$

Therefore,

$$||x_N|| \leq k_0 ||x_0|| + \sum_{n=0}^{N-1} k_0 h_n (k_1 ||x_n|| + k_2) \quad (32)$$

Consider the case $k_1 > 0$ first. We show (30) by induction. It is evidently true for $N = 0$ since $k_0 \geq ||\hat{S}_n^n|| = 1$. Assuming its validity for $n < N$, we substitute in (32) to get

$$\begin{aligned}
||\underline{x}_N|| &\leq k_0 ||\underline{x}_0|| + \sum_{n=0}^{N-1} k_0 k_1 h_n [k_0 ||\underline{x}_0|| + \frac{k_2}{k_1}] e^{k_0 k_1 (t_n - t_0)} \\
&= [k_0 ||\underline{x}_0|| + \frac{k_2}{k_1}] [1 + \sum_{n=0}^{N-1} k_0 k_1 h_n e^{k_0 k_1 (t_n - t_0)}] - \frac{k_2}{k_1} \\
&\leq [k_0 ||\underline{x}_0|| + \frac{k_2}{k_1}] [1 + \sum_{n=0}^{N-1} (e^{k_0 k_1 h_{n-1}}) e^{k_0 k_1 (t_n - t_0)}] - \frac{k_2}{k_1} \\
&= [k_0 ||\underline{x}_0|| + \frac{k_2}{k_1}] e^{k_0 k_1 (t_N - t_0)} - \frac{k_2}{k_1}
\end{aligned}$$

This completes the inductive proof of (30). If $k_1 = 0$, equation (31) follows from (32) directly.

Q.E.D.

The convergence theorem is now straightforward.

Theorem 2 If a method satisfies the stability condition with respect to a step selection scheme θ , if α, α^*, β and β^* or A_n and $\underline{\ell}_n$ are uniformly bounded, and if the truncation error $\underline{d}_n$ satisfies

$$||\underline{d}_n|| \leq h_n e(h) \quad \forall n$$

where $e(h) \rightarrow 0$ as $h \rightarrow 0$, then the method is stable and converges with respect to θ , and there exist constants k_3 and k_4 such that the global error $\underline{\varepsilon}_N$ satisfies

$$||\underline{\varepsilon}_N|| \leq k_3 ||\underline{\varepsilon}_0|| + k_4 e(h) \text{ for } 0 \leq t_N \leq T$$

Notes: 1. The condition on $\underline{d}_n$ is a consistency condition. If the method has order r , then $e(h) = O(h^r)$. However, a more useful way of looking at this result is that if we control the step size h_n so that the

truncation error is bounded by ϵh_n , the global error is bounded by $k_3 ||\epsilon_0|| + k_4 \epsilon$, so that changes in the truncation error control are proportionately reflected in the global error bound.

2. The condition that α , α^* , β and β^* or A_n and ℓ_n be uniformly bounded is also a form of consistency. It rules out the following type of case:

The explicit formula

$$y_{n+1} = \frac{1}{h_{n-2} - h_{n-1}} \left[\frac{(h_n + h_{n-1})^2 - h_{n-2}^2}{h_{n-2} + h_{n-1}} y_n + \frac{(h_n + h_{n-1})^2 - h_{n-1}^2}{h_{n-2} + h_{n-1}} y_{n-2} - \frac{h_n (h_n + h_{n-1} + h_{n-2})}{h_{n-2}} y'_{n-1} \right]$$

is second order provided that $h_{n-2} \neq h_{n-1}$. However, as $h_{n-2} \rightarrow h_{n-1}$, the coefficients and the error term blow up. If we used this formula with a step selection scheme which kept h_{n-2}/h_{n-1} bounded away from one, we would get convergence if the method was stable, but if the step selection scheme let $h_{n-1}/h_{n-2} \rightarrow 1$ as $h \rightarrow 0$, we may not get convergence.

Proof From (20) and (24) the error equation for a method is given by

$$\epsilon_{n+1} = S_n \epsilon_n + \frac{d_n}{h_n} \quad (33)$$

$$= \hat{S}_n \epsilon_n + h_n \tilde{S}_n \epsilon_n + \frac{d_n}{h_n}$$

Consistency and Lemma 1 imply that $||\tilde{S}_n|| \leq k_1$ for some $k_1 < \infty$ provided that $h_n \leq u$. Lemma 2 can be applied to (33) with $k_2 = e(h)$ since the stability hypothesis guarantees the existence of a constant

k_0 such that $||\hat{S}_n^m|| \leq k_0$.

Therefore,

$$\begin{aligned} ||\underline{\varepsilon}_N|| &\leq [k_0 ||\underline{\varepsilon}_0|| + \frac{e(h)}{k_1}] e^{k_0 k_1 (t_N - t_0)} - \frac{e(h)}{k_1} \\ &\leq k_3 ||\underline{\varepsilon}_0|| + k_4 e(h) \quad \forall 0 \leq t_N \leq T \end{aligned} \quad (34)$$

Hence, $||\underline{\varepsilon}_N|| \rightarrow 0$ if $||\underline{\varepsilon}_0||$ and $e(h) \rightarrow 0$ and we have convergence.

The difference between two numerical solutions satisfies

$$\underline{y}_{n+1} - \hat{y}_{n+1} = S_n(y_n - \hat{y}_n)$$

The same analysis (with $e(h) = 0$) yields

$$||\underline{y}_m - \hat{y}_m|| \leq k_3 ||\underline{y}_n - \hat{y}_n||$$

Therefore, the method is stable.

Q.E.D.

The difficult problem left is to verify the stability condition for various methods and step selection schemes. First, we will deal with the Adams methods in Theorems 3, 4, 5, and 6. Theorem 5 is an extension of the result of Piotrowski [5].

Theorem 3 A_n and $\underline{\ell}_n$ are bounded for the interpolation technique using a step selection scheme Θ .

Proof For the interpolation technique $\underline{\ell}_n = \underline{\ell}$ is independent of n , and is, therefore, bounded. $A_n = A C_n$ where C_n , given by equation (12), is bounded by Δ^{-2k+1} from definition 1. Therefore, A_n is bounded.

Q.E.D.

Theorem 4 α, α^*, β and β^* are bounded for the variable step Adams method using a step selection scheme θ .

Proof The variable step Adams method has the form

$$y_{n+1} = y_n + \beta_{0,n}^* h_n y'_{n+1} + \dots + \beta_{k,n}^* h_{n-k} y'_{n-k+1} \quad (35)$$

where the $\beta_{i,n}^*$ are determined from the requirement that the method have order $k+1$. (For an explicit method, $\beta_{0,n}^* = 0$ and the order is k .) A necessary and sufficient condition for the order to be $k+1$ is that (35) be exact for the functions $y(t) = (t - t_{n+1})^r$, $0 \leq r \leq k+1$. This leads to a system of $k+1$ linear equations for the $\beta_{i,n}^*$, namely

$$\begin{aligned} \sum_{i=0}^k \beta_{i,n}^* h_{n-i} r(t_{n+1-i} - t_{n+1})^{r-1} \\ = -(-h_n)^r, \quad r = 1, 2, \dots, k+1 \end{aligned} \quad (36)$$

Define

$$\begin{aligned} \omega_i &= \frac{t_{n+1-i} - t_{n+1}}{h} \\ &= -\frac{h_n + h_{n-1} + \dots + h_{n+1-i}}{h} \end{aligned}$$

From definition 1

$$\Delta \leq |\omega_i| < i, \quad i \geq 1 \quad (37)$$

$$0 = \omega_0$$

and

$$|\omega_i - \omega_j| \geq \Delta \quad \text{if } i \neq j \quad (38)$$

Divide (36) by rh^r to get

$$\sum_{i=0}^k \beta_{i,n}^* (\omega_i - \omega_{i+1}) \omega_i^{r-1} = -\frac{\omega_1^r}{r} \quad (39)$$

The right hand side of (39) and the coefficients of $\beta_{i,n}^*$ on the left hand side are bounded uniformly in n by (37). The determinant D of the coefficients is a simple multiple of the Vandermonde determinant, namely

$$D = \left[\prod_{k \geq i > j \geq 0} (\omega_i - \omega_j) \right] \left[\prod_{i=0}^k (\omega_i - \omega_{i+1}) \right] \quad (40)$$

By (38) this is bounded away from zero, hence the solution of (39) is bounded uniformly in n .

Q.E.D.

Theorem 5 *The variable step Adams-Bashforth (AB), Adams-Bashforth-Moulton predictor-corrector (ABM), and Adams-Moulton (AM) methods are stable and convergent for any step selection scheme.*

Proof The vector of saved values $\underline{y}_n$ is $[y_n, hy'_n, \dots, hy'_{n-k+1}]^T$ so the matrix A_n is given by

$$\begin{bmatrix} 1 & \beta_{1,n} & \beta_{2,n} & \dots & \beta_{k,n} \\ 0 & \delta_{1,n} & \delta_{2,n} & \dots & \delta_{k,n} \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

while $\underline{\ell}_n$ is $[0, 1, 0, \dots, 0]^T$ for AB and $[\beta_{0,n}^*, 1, 0, \dots, 0]^T$ for ABM and AM.

Hence, from equation (23),

$$\hat{S}_n = (I - \frac{\ell_n}{2} e_2^T)^M A_n$$

$$= \begin{bmatrix} 1 & \eta_{1,n} & \eta_{2,n} & \dots & \eta_{k,n} \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (41)$$

where

$$\eta_{i,n} = \begin{cases} \beta_{i,n} & \text{for AB} \\ \beta_{i,n}^* & \text{for ABM and AM} \end{cases}$$

From Theorem 4, $\eta_{i,n}$ are uniformly bounded, therefore, $\hat{S}_n$ is uniformly bounded. Hence the $||\hat{S}_n^{n+m}||$ are uniformly bounded for $0 \leq m \leq k$. From the structure of $\hat{S}_n$, we see that for $m > k$

$$\hat{S}_n^{n+m} = \hat{S}_n^{n+k} = \begin{bmatrix} 1 & x_1 & x_2 & \dots & x_k \\ & & & & \\ & & 0 & & \\ & & & & \end{bmatrix} \quad (42)$$

Hence the $||\hat{S}_n^{n+m}||$ are uniformly bounded for all $m \geq n$ and the method satisfies the stability condition. From Theorem 4, A_n and ℓ_n are uniformly bounded. Finally, we can expand $y(t_{n-i})$ and $hy'_{n-i-1}(t_{n-i})$ by Taylor's series about t_{n+1} with a remainder involving $h^r y^{(r)}(\xi)$.

If the order is greater than $r-2$, then all but the remainder terms will vanish, and we can bound $||\underline{d}_n||$ by $h_n(h^{r-1} C_r \text{Max}|y^{(r)}|)$.

Thus, the three conditions of Theorem 2 are satisfied.

Q.E.D.

Theorem 6 *If the step selection scheme is such that at least k steps of constant size are taken between step changes, then the k -step interpolation AB, ABM, and AM methods are stable and convergent.*

Proof Once again, we show that the hypothesis of Theorem 2 are satisfied. We will use the Nordsieck vector to express the method as a $k+1$ value method. Then the matrix S_n is given by

$$\begin{aligned}\hat{S}_n &= [I - \underline{\ell} \underline{e}_2^T]^M A C_n \\ &= R C_n\end{aligned}\tag{43}$$

The matrix R has the important properties that the simple eigenvalue one corresponds to the right eigenvector $\underline{e}_1$, and that all other eigenvalues are zero (Gear [2], pp. 138-140). Consequently, $R^k \underline{x} = \mu_x \underline{e}_1$ where $|\mu_x|/||\underline{x}||$ is uniformly bounded for all $\underline{x}$ (Tu [6], p. 20). The eigenvalue one of C_n corresponds to the eigenvector $\underline{e}_1$. When $h_n = h_{n-1}$, $C_n = I$. From the hypothesis of the theorem, it is not possible for $C_i \neq I \neq C_j$ if $|i-j| < k$. Consequently, the product

$$\hat{S}_n^m = R C_{m-1} R C_{m-2} \dots R C_n$$

either contains less than $2k+2$ matrices, or can be reduced to

$$\hat{S}_n^m = R C_{m-1} \dots R C_q R^k C_{q-k} R^{q-k-n}$$

where

$$0 \leq q-k-n < k$$

Since the ratio of adjacent steps is bounded by Δ^{-1} , $||C_m||$ is uniformly bounded. Hence, $||C_{q-k} R^{q-k-n}|| \leq k_5$ uniformly. Therefore, if $\hat{S}_n^m$ contains $2k+2$ or more matrices

$$\begin{aligned}\hat{S}_n^m \underline{z} &= R C_{m-1} \dots R C_q R^k C_{q-k} R^{q-k-n} \underline{z} \\ &= R C_{m-1} \dots R C_q R^k \underline{x}\end{aligned}$$

where $||\underline{x}|| \leq k_5 ||\underline{z}||$.

Hence,

$$\begin{aligned}\hat{S}_n^m \underline{z} &= R C_{m-1} \dots R C_q \mu_x \underline{e}_1 \\ &= \mu_x \underline{e}_1\end{aligned}$$

so that

$$\begin{aligned}||\hat{S}_n^m \underline{z}|| &= |\mu_x| \\ &\leq k_6 ||\underline{x}|| \quad (\text{as } \mu_x / ||\underline{x}|| \text{ is bounded}) \\ &\leq k_5 k_6 ||\underline{z}||\end{aligned}$$

Therefore,

$$||\hat{S}_n^m|| \leq k_5 k_6$$

If $\hat{S}_n^m$ contains less than $2k+2$ matrices, $||\hat{S}_n^m||$ is bounded by the finite product of the norms of each matrix, hence the method satisfies the stability condition.

Theorem 3 shows that A_n and $\underline{l}_n$ are bounded. $\underline{d}_n$ is bounded by the same argument as used in Theorem 5. Hence Theorem 2 proves convergence and stability.

Q.E.D.

Tu [6] points out that the interpolation form of the Adams formula is not necessarily stable under any step selection scheme. He shows that for the three step ABM formula with

$$h_{2i} = \omega h_{2i+1} = h_{2i+2} = \omega h_{2i+3} = \dots$$

the matrix $\hat{S}_{2i}^{2i+2m}$ has an eigenvalue of $[(\omega-1)^2/4\omega]^m$. For sufficiently large ω this is unbounded. Examples of instability when $\omega = 10$ are given in that thesis along with examples showing that the variable step technique is stable for the same step selection scheme. This case points out that the variable step technique is more stable in some cases.

Another example in Tu's report substantiates Brayton, et.al.'s claim that the variable step technique is more stable. The problem $y' = -y$ with $h_{2i} = .05$ and $h_{2i+1} = .005$ was integrated using a three step backward differentiation formula. It was apparently stable for the variable step technique and apparently unstable for the interpolation technique. It is hypothesized that the variable step method can also cause instabilities, but no examples have yet been found.

In practice, we do not vary the step in an extreme way. When the step changes are controlled to be small, both techniques are stable as shown below.

Theorem 7 *If a method satisfies the stability condition for fixed steps, it also does so with respect to a step selection scheme that produces step size changes small in the sense that*

$$\frac{h_{n+1}}{h_n} = 1 + O(h) \text{ as } h \rightarrow 0 \quad (44)$$

Note: If $\theta(t, h) \equiv \theta(t)$ and θ is a boundedly differentiable function of t , then the hypotheses of the theorem are satisfied since

$$\begin{aligned} \frac{h_{n+1}}{h_n} &= \frac{\theta(t_{n+1})}{\theta(t_n)} = 1 + h_n \frac{\theta'(\xi_n)}{\theta(t_n)} \\ &= 1 + h\theta'(\xi_n) \\ &= 1 + O(h) \end{aligned}$$

Proof We will use the hypotheses of the theorem to decompose $\hat{S}_n$ into the form

$$\hat{S}_n = \hat{S} + h_n \bar{S}_n \quad (45)$$

where $\hat{S}$ is the form of $\hat{S}_n$ when constant steps are used and $\bar{S}_n$ can be bounded. (This step will be delayed until Lemma 3 below.) Then we can examine $||\hat{S}_n^m||$ as follows. For fixed n , set

$$\underline{x}_m = \hat{S}_n^m \underline{x}_n \quad m \geq n \quad (46)$$

Then

$$\begin{aligned} \underline{x}_{m+1} &= \hat{S}_m \underline{x}_m \\ &= \hat{S} \underline{x}_m + h_m \bar{S}_m \underline{x}_m \end{aligned} \quad (47)$$

If the constant step method satisfies the stability condition, then

$||\hat{S}^{m-n}|| \leq k_0$. Lemma 3 will show that $||\bar{S}_n|| \leq k_1$. Therefore, Lemma 2 can be applied to (47) with $k_2 = 0$. Hence,

$$||\underline{x}_m|| = ||\hat{S}_n^m \underline{x}_n|| \leq k_0 e^{k_0 k_1^T} ||\underline{x}_n||$$

Hence,

$$||\hat{S}_n^m|| \leq k_0 e^{k_0 k_1^T}$$

so the method satisfies the stability condition.

Q.E.D.

Lemma 3 If a step selection scheme satisfies the hypothesis of Theorem 7, $\hat{S}_n$ can be written as $\hat{S} + h_n \bar{S}_n$ where $\hat{S}$ is independent of h and n , and $||\bar{S}_n||$ is uniformly bounded.

ProofPart I: Interpolation Technique

From (9) and (23)

$$\hat{S}_n = (I - \underline{\ell}_n \underline{e}_2^T)^M A C_n$$

where $\underline{\ell}_n$ is independent of n . Now

$$\begin{aligned} C_n &= \text{diag}(1, h_n/h_{n-1}, (h_n/h_{n-1})^2, \dots) \\ &= I + \text{diag}(0, h_n/h_{n-1} - 1, (h_n/h_{n-1})^2 - 1, \dots) \\ &= I + O(h) \\ &= I + O(h_n) \end{aligned}$$

since $h_n/h_{n-1} = 1 + O(h)$ and $h \leq h_n \Delta^{-1}$.

Therefore,

$$\hat{S}_n = (I - \underline{\ell}_n \underline{e}_2^T)^M A(I + O(h_n)) = \hat{S}[I + O(h_n)]$$

Hence,

$$\bar{S}_n = \hat{S} O(h_n)/h_n$$

is uniformly bounded.

Part II: Variable Step Technique

It is shown in Tu [6], p. 52, that the coefficients α and β have the form

$$\alpha_{i,n} = \alpha_i + R_{\alpha_i}(v_n, \dots, v_{n-k+2})$$

$$\beta_{i,n} = \beta_i + R_{\beta_i}(v_n, \dots, v_{n-k+2})$$

where

$$v_i = \frac{h_i - h_{i-1}}{h_{i-1}}$$

where the R 's are rational functions that vanish at $v_n = v_{n-1} = \dots = v_{n-k+2} = 0$ (when the step sizes are equal), and where the α_i and β_i are the corresponding fixed step coefficients. (This result follows directly from the nonsingular system of equations that determine α and β .) From the hypothesis of Theorem 7, $v_i = O(h)$, hence $R = O(h) = O(h_n)$. Therefore, we can express $\hat{S}_n$, which involves the coefficients α , β , α^* and β^* , as $\hat{S} + h_n \bar{S}_n$, where $\hat{S}$ is the matrix obtained by setting the $\alpha_{i,n}$ equal to α_i , etc., and $\bar{S}_n$ includes the terms R divided by h . These can be uniformly bounded.

Q.E.D.

5. PRACTICAL IMPLICATIONS

It can be pointed out that a common step selection scheme satisfies the hypotheses of Theorem 7. If the step h_n is controlled so that the local error estimate is equal to ϵ (or ϵh_n), and if the truncation error estimate is $\phi(t_n) h_n^{r+1} + O(h_n^{r+2})$ where $\phi(t)$ is the principle error function for the method, we have

$$\phi(t_n) h_n^{r+1} + O(h_n^{r+2}) = \epsilon \text{ (or } \epsilon h_n \text{)}$$

Hence,

$$h_n = \left(\frac{\epsilon}{\phi(t_n)} \right)^{1/q} + O(h_n^2) \text{ where } q = r+1 \text{ or } r,$$

provided that $\phi(t_n)$ is not zero. If $\phi(t)$ is bounded away from zero and $\phi'(t)$ is bounded above, then

$$\begin{aligned} \frac{h_{n+1}}{h_n} &= \left(\frac{\phi(t_n)}{\phi(t_{n+1})} \right)^{1/q} + O(h_n^2) \\ &= 1 + O(h_n) \end{aligned}$$

and $h_{\min}/h_{\max} \geq \Delta > 0$. Consequently, we can expect either method to be stable if the fixed step method is stable.

However, in practice we do not send h to the limit and we often do not bother to adjust the step if the change is small. Therefore, there is reason to prefer the variable step approach. Brayton, et.al. [1] point out that the amount of work in evaluating $A_n y_n$ is less in the variable step approach than in the interpolation/Nordsieck approach, but that the coefficients in A_n and $\underline{y}_n$ must be recomputed for each step. For a large system of equations, this is not important as A_n is the same for each equation in the system. For a few equations this is a serious consideration, which leads us to recommend variable step methods for large systems of equations and interpolation methods with the Nordsieck vector for small systems.

LIST OF REFERENCES

- [1] Brayton, R. K., Gustavson, F. G., and Hachtel, R. D. "A New Efficient Algorithm for Solving Differential-Algebraic Systems Using Implicit Backward Difference Formulas," Proceedings of the IEEE, 60, #1, pp. 98-108, 1972.
- [2] Gear, C. W. NUMERICAL INITIAL VALUE PROBLEMS IN ORDINARY DIFFERENTIAL EQUATIONS, Prentice-Hall, Inc., New Jersey, 1971.
- [3] Henrici, P. DISCRETE VARIABLE METHODS FOR ORDINARY DIFFERENTIAL EQUATIONS, Wiley, New York, 1962.
- [4] Nordsieck, A. "On the Numerical Integration of Ordinary Differential Equations," Math. Comp., 16, pp. 22-49, 1962.
- [5] Piotrowski, P. "Stability, Consistency and Convergence of Variable K-Step Methods," Conference on the Numerical Solution of Differential Equations, 109, Springer, Berlin, 1969.
- [6] Tu, K. W. "Stability and Convergence of General Multistep and Multivalued Methods with Variable Step Size," Department of Computer Science Report 526, University of Illinois at Urbana-Champaign, July 1972 (Ph.D. Thesis).

U. S. ATOMIC ENERGY COMMISSION
UNIVERSITY-TYPE CONTRACTOR'S RECOMMENDATION FOR
DISPOSITION OF SCIENTIFIC AND TECHNICAL DOCUMENT

(See Instructions on Reverse Side)

1. AEC REPORT NO.

COO-1469-0220

2. TITLE

THE EFFECT OF VARIABLE MESH SIZE ON
THE STABILITY OF MULTISTEP METHODS

3. TYPE OF DOCUMENT (Check one):

- ☒ a. Scientific and technical report
☐ b. Conference paper not to be published in a journal:
 Title of conference _____
 Date of conference _____
 Exact location of conference _____
 Sponsoring organization _____
☐ c. Other (Specify) _____

4. RECOMMENDED ANNOUNCEMENT AND DISTRIBUTION (Check one):

- ☒ a. AEC's normal announcement and distribution procedures may be followed.
☐ b. Make available only within AEC and to AEC contractors and other U.S. Government agencies and their contractors.
☐ c. Make no announcement or distribution.

5. REASON FOR RECOMMENDED RESTRICTIONS:

6. SUBMITTED BY: NAME AND POSITION (Please print or type)

C. W. Gear
Professor and Principal Investigator

Organization

Department of Computer Science
University of Illinois
Urbana, Illinois 61801

Signature

Charles W. Gear

Date

April 1973

FOR AEC USE ONLY

7. AEC CONTRACT ADMINISTRATOR'S COMMENTS, IF ANY, ON ABOVE ANNOUNCEMENT AND DISTRIBUTION RECOMMENDATION:

8. PATENT CLEARANCE:

- ☐ a. AEC patent clearance has been granted by responsible AEC patent group.
☐ b. Report has been sent to responsible AEC patent group for clearance.
☐ c. Patent clearance not required.

BIBLIOGRAPHIC DATA SHEET		1. Report No. UIUCDCS-R-73-570	2.	3. Recipient's Accession No.
4. Title and Subtitle THE EFFECT OF VARIABLE MESH SIZE ON THE STABILITY OF MULTISTEP METHODS			5. Report Date April 1973	
			6.	
7. Author(s) C. W. Gear, K. W. Tu			8. Performing Organization Rept. No.	
9. Performing Organization Name and Address Department of Computer Science University of Illinois Urbana, Illinois 61801			10. Project/Task/Work Unit No.	
			11. Contract/Grant No. US AEC AT(11-1)1469	
12. Sponsoring Organization Name and Address US AEC Chicago Operations Office 9800 South Cass Avenue Argonne, Illinois 60439			13. Type of Report & Period Covered technical	
			14.	
15. Supplementary Notes				
16. Abstracts The effects of two different techniques for implementing variable mesh sizes in multistep formulas are investigated. It is proved that one is more stable than the other for some cases, but that both are stable when the step changes are small. The practical implications of these results are discussed.				
17. Key Words and Document Analysis. 17a. Descriptors interpolation variable step				
17b. Identifiers/Open-Ended Terms				
17c. COSATI Field/Group				
18. Availability Statement unlimited			19. Security Class (This Report) UNCLASSIFIED	
			20. Security Class (This Page) UNCLASSIFIED	
			21. No. of Pages 35	
			22. Price	

OCT 29 1973



UNIVERSITY OF ILLINOIS-URBANA
510.84 IL6R no. C002 no.565-570(1973
Two upper bounds for the weighted peth I



3 0112 088400624